
Proceedings Paper:
Kampanos, Georgios and Shahandashti, Siamak F. orcid.org/0000-0002-5284-6847 (Accepted: 2021) Accept All : The Landscape of Cookie Banners in Greece and the UK. In: International Conference on ICT Systems Security and Privacy Protection, Proceedings. International Conference on ICT Systems Security and Privacy Protection, 22-24 Jun 2021 Springer , NOR , p. 213. (In Press)

https://doi.org/10.1007/978-3-030-78120-0_14

Accept All: The Landscape of Cookie Banners in Greece and the UK*

Georgios Kampanos and Siamak F. Shahandashti

University of York, UK
kampanosg@outlook.com, siamak.shahandashti@york.ac.uk

Abstract. Cookie banners are devices implemented by websites to allow users to manage their privacy settings with respect to the use of cookies. They are part of a user's daily web browsing experience since legislation in Europe requires websites to show such notices. In this paper, we carry out a large-scale study of more than 17,000 websites including more than 7,500 cookie banners in Greece and the UK to determine compliance and tracking transparency levels. Our analysis shows that although more than 60% of websites store third-party cookies in both countries, only less than 50% show a cookie notice and hence a substantial proportion do not comply with the law even at the very basic level. We find only a small proportion of the surveyed websites providing a direct opt-out option, with an overwhelming majority either nudging users towards privacy-intrusive choices or making cookie rejection much harder than consent. Our results differ significantly in some cases from previous smaller-scale studies and hence underline the importance of large-scale studies for a better understanding of the big picture in cookie practices.

Keywords: Cookie Banners · Privacy Options · Web Measurement · Dark Patterns · GDPR · Data Protection Act · User Tracking.

1 Introduction

Websites implement cookie banners to allow users to either consent to or reject third-party cookie tracking and manage their privacy settings. After tighter legislation came into force, namely EU's General Data Protection Regulation (GDPR) and UK's Data Protection Act 2018 (DPA), more and more websites have adopted such notices, making cookie banners a part of users' everyday life.

In theory, cookie banners (a.k.a. cookie notices) exist to empower users by informing them about tracking activity and allowing them to opt out if they wish to. However, real-world implementations of cookie banners appear to be a nuisance more than anything else [9]. Many websites design their notices to make opting out extremely hard, or remove the option completely as previous studies found [13]. Furthermore, pre-selected options that nudge users towards privacy-intrusive choices are rife and significantly impact user behaviour [14]. Both EU

* This article is accepted to and to appear in the proceedings of IFIP SEC 2021.

and UK regulators have clearly identified such practices non-compliant with the GDPR and the DPA, including consent not being explicit and cookie rejection not being as easy as acceptance (See e.g. the EU's 2002 ePrivacy Directive [7], the 2020 European Data Protection Board guidelines [6], and discussions by Nouwens et al. [13]). Yet, even if users manage to navigate around the maze of options and select their privacy settings, their choices are more likely to be ignored entirely as a study of Consent Management Providers (CMPs) deployed in European websites observed [11]. Worryingly, we have seen such "dark patterns" employed by big-tech, such as Facebook [12].

The insight we have about the cookie banner landscapes and how they have changed as a result of legislation is mainly based on the analyses carried out on samples of high-traffic websites. Although such studies provide valuable information on how popular websites implement cookie banners, a natural question to ask is how well such observations generalise if a more comprehensive sample including lower-traffic websites is analysed. In this work, we aim to take a step towards investigating this question in the UK and Greece web landscapes.

We set out to establish the types of cookie banners with which users have to interact on a daily basis. Moreover, we will explore the distribution and availability of choices provided to users through cookie banner implementations. Using purpose-built software and with the aid of OpenWPM [4], we collected, categorised and analysed more than 7,500 cookie banners from more than 17,000 websites across Greece and the UK. We discuss our findings which interestingly in some cases substantially differ with previous results in the literature. Our results therefore is a step towards developing a more clear and comprehensive understanding of the cookie banner landscape in the two countries.

We consider Greece and the UK because of our familiarity with the respective languages and our hope that the comparison between the two provides interesting insight. On the one hand, websites in both countries adhere to very similar data protection laws. On the other hand however, the two countries vastly differ in their population size and their citizens' use of internet services [5].

2 Related Work

Studies in this area have mainly focused on the prevalence of cookie banners, the type of privacy options they offer, and whether they comply with the law.

In 2018, The Norwegian Consumer Council reviewed whether user interfaces of cookie notices and privacy settings provided by Google, Facebook and Microsoft Windows 10 discourage users from making privacy-aware choices [12]. They found that all three companies offer default settings that are considered privacy intrusive, and that the cookie notices contain misleading wording while privacy-friendly options require multiple steps to find. They noted that Google and Facebook "threaten users with loss of functionality or deletion of the user account" unless they agree to those privacy-intrusive settings.

A number of studies in this area focus on providing a big picture across the world or Europe. Habib et al. conducted a 150-website analysis in 2018–19 and

found that privacy options are frequent within their sample with 89% websites with targeted advertising offering a way to opt-out [8]. However, they observed that, when visited from the US, only 28 out of 150 websites they considered displayed a cookie banner with only 5 of them offering a means to opt out. Degeling et al.'s study of 6,759 websites across the EU found adoption of cookie banners across the EU go up from 46.1% before the GDPR to 62.1% afterwards [2]. Utz et al. carried a manual inspection of 1,000 popular websites in the EU and observed that 27.8% provide no options, 68.0% allow confirmation only, while only 3.2% give a binary accept/reject choice [14]. Another study of top 100 websites in each EU country by van Eijk et al. found 52% of UK and 29% of Greek websites implementing a cookie banner [3].

Two recent studies have looked at whether cookie banners provided by Content Management Platforms (CMPs) adhere to EU regulations. In a study published in 2019, Matte, Bielova and Santos surveyed 1,427 European websites from which they found that 141 websites registered an affirmative consent before the user had performed any actions and 38 websites offered no "opt-out" option at all [11]. The authors observed that at least 50% of the websites in their dataset had pre-selected privacy options and at least 27 websites did not respect the user's choice even though they declined to be tracked by cookies. In a study published in 2020 considering 680 websites, Nouwens et al. found that 32% of them assumed "implicit consent" (agreeing without having any other option) [13], which make those websites not-compliant with GDPR. They also found that only 13% of websites had a "reject all" button which almost always required additional clicks to be seen by a user.

Taking a closer look at the sample sizes of the studies that focus on providing a big picture, we have Habib et al.'s study of 150 websites worldwide (50 from each group of high, middle, and low popularity) [8], Degeling et al.'s sample of 6,759 websites including 463 UK and 443 Greek ones [2], Utz et al.'s 1,000 randomly chosen from 500 top-ranking in each EU country [14], and van Eijk et al.'s sample of top 100 in each EU country [3]. Similarly for the studies that limit their attention to websites with CMP-provided banners, these include Nouwens et al.'s study of 680 such websites in the UK [13] and Matte, Bielova, and Santos's investigation of 1,426 websites including 149 `.uk` and 53 `.gr` websites [11]. Although such studies provide valuable insight, none goes beyond 700 websites in the two countries we consider. This is understandable due to the complexity of automating such studies and that the goal of the said studies was to focus on a global view or on CMPs, and not on the comprehensive landscape in specific countries. This opens a natural question whether similar trends can be seen if the scale of the sample sizes considered are increased. Indeed, it is not clear whether characteristics observed in high-traffic websites remain similar if low-traffic ones are considered. To answer this question, we focused on two specific countries: the UK and Greece, but scaled up the sample size nearly ten-fold by automating our scraping and analysis, allowing us to expand our research to more than 14,000 UK and 3,000 Greek websites.

While we developed purpose-built software, this study relied heavily upon existing software. We extended OpenWPM, an open-source web privacy measurement tool developed by Englehardt and Narayanan in 2016 to scrape and collect data [4]. OpenWPM allows researchers to detect and measure the use of third-party cookies (TPs), cookie synchronisation, as well as fingerprinting techniques. We modified OpenWPM to be able to recognise and store the website's cookie banner if one exists.

3 Research Questions, Methodology, and Implementation

In light of the results of the previous works, we aim to investigate the following research questions through a large-scale study of Greek and UK domains:

RQ1: What is the prevalence of cookie banners across the board when less popular websites in Greece and the UK are also considered?
RQ2: How does the distribution of options offered in cookie banners look like and what proportion of websites provide a direct cookie rejection option?
RQ3: What proportion of websites employ implicit consent?
RQ4: What proportion of cookie banners allows their users to manage their privacy settings and control which vendors track them?
RQ5: How do the countries compare in terms of the privacy options offered by cookie banners?

Our data collection included three steps. We first built a comprehensive set of functioning websites to analyse and extract their cookie banners. Then we crawled the identified websites and collected relevant data such as the source code of the cookie notices and screenshots of the webpages. Finally, we sanitised and structured the collected data into a data structure that facilitates analysis. In the following, we explain these steps in more detail. The code developed for this study and referred to throughout the paper is publicly available at the following repository: `https://github.com/kampanosg/i-like-cookies`.

3.1 Building the Target List

The first step is to identify websites to be analysed in this study. Using the Tranco top sites ranking [10], popular websites for the two countries were identified based on their Top Level Domains (TLDs): `.uk` and `.gr`. We decided to augment the TLD-based lists with other curated country-specific lists since many websites do not use the TLD of their country of origin, e.g., British Airways uses `.com`.

Ethical Considerations. Not all websites allow crawling and many explicitly state that they only allow "personal use" of their services and content to their visitors. To respect such restrictions imposed by the websites, we developed two parsers to identify and exclude such websites from our automated crawl:

1. A Robots Exclusion Standard parser (`step1b_checker_robots.py`) that verifies whether websites allow crawling by reading their `robots.txt` file;

2. A Terms of Service (ToS) parser (`step1c_checker_tos.py`) that makes best effort to find exclusionary terms, e.g. "for personal use only", to comply with the ToS of each website.

3.2 Collecting Cookie Banners

The second step is to effectively identify and collect the cookie banners on the compiled set of websites. We achieve this by taking advantage of the "I don't care about cookies" (IDCAC) list (`www.i-dont-care-about-cookies.eu`), which provides an extensive selection of standard CSS selectors that cookie banners use. We parse these selectors and add them to a database. Furthermore, during testing, we identified and added 64 additional selectors to IDCAC.

After setting up the cookie selectors database, OpenWPM uses it to identify the cookie banners within the visited websites. We extended OpenWPM to detect cookie banners within a given website. For each website, we check whether it contains a CSS selector from the cached IDCAC list, using Selenium (`www.selenium.dev`), which allows for searching the HTML code of the website. When OpenWPM identifies a cookie banner, we perform additional analysis to make sure it is not a false positive, briefly by first ensuring that Selenium returned a valid HTML of a reasonable length, and then verifying that the returned HTML contains the terms "cookie" or "cookies".

When OpenWPM successfully identifies a cookie banner, it is stored in a database for further analysis. Combining the cached cookie selectors and Selenium's efficient Document Object Model (DOM) search enables the cookie banner extension to be efficient and robust.

3.3 Classifying and Normalising the Data

To our knowledge, no standard exists for cookie banners, and therefore, every website has a different implementation for their notices. Thus, the HTML code and the options they provide can be drastically different from website to website. Such complexity can make the data analysis difficult. Thus, before performing any research on the data, we transformed them into a consistent data structure.

First, we sanitised the collected data to identify and then classify the privacy options within the collected cookie banners. We identified four privacy option categories: *Affirmative*, *Negative*, *Informational* and *Managerial* as defined in Table 1. We developed these four categories by manually inspecting a random sample of the collected data during testing, further informed by our own experience with cookie banners in the wild. Using these categories allows us to classify the cookie banners by the types of options they provide and hence better understand user choices. Examples of cookie banners providing different combinations of these options can be seen in Fig. 1.

Manual inspection of the privacy options was necessary to account for local nuances. For example, although the noun $\alpha\pi o\delta o\chi\eta$ (lit. acceptance) was the most popular Affirmative call to action, a large number of banners used the verb $\delta\epsilon\chi o\mu\alpha\iota$ (I accept) indicating the same. We developed a comprehensive list of

Table 1: The developed cookie banner options categories.

Category	Description
Affirmative	Options that prompt users to accept the use of cookies, e.g. “accept”, “agree”, “allow”, and “OK”.
Negative	Options that allow users to opt-out from cookie tracking, e.g. “decline”, “reject”, “disagree”, and “no”.
Informational	Options that take users to informational pages, e.g. “Privacy Policy”, “learn/see more”, and “see/show details”.
Managerial	Options that allow users to opt in/out of specific trackers, e.g. “manage”, “settings”, and “vendors/partners”.

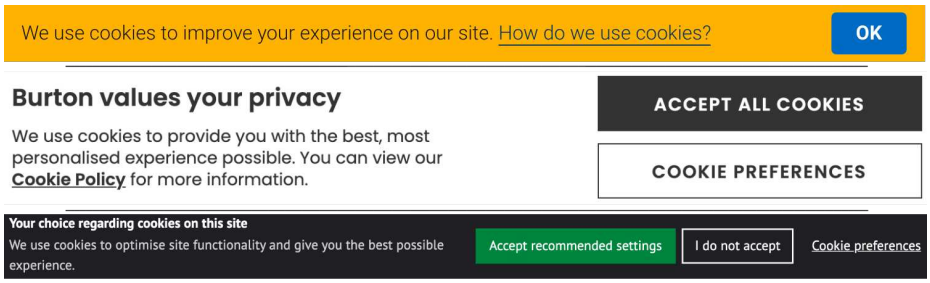

Fig. 1: Three examples of cookie banners with different privacy options. Top: Affirmative and Informational, Middle: Affirmative, Managerial, and Informational, Bottom: Affirmative, Negative, and Managerial.

such variations and wrote a Python script (`step3a_parse_cookie_banners.py`) to categorise cookie banners.

After we categorised the privacy options, we transformed the collected data into a consistent data structure that allows for efficient querying. More specifically, we converted the arbitrary HTML form of the collected cookie notices into JSON to facilitate both manual and automated analyses. Fig. 2 depicts an example banner before and after categorisation and normalisation.

4 Data and Results

In this section we specify the dataset and discuss our findings.

4.1 The Collected Dataset

Websites. The Tranco list contains a total of 1 M websites. From there, 3,446 are `.gr` websites and 18,768 are `.uk` ones. The additional country-specific lists provided an additional 674 websites: 40 additional Greek websites from TopGR (`https://topgr.gr`) and 634 additional UK websites from Kadaza (`www.kadaza.co.uk`) and Finder (`www.finder.com/uk`). Furthermore, we removed 125 Greek

```
<div class="..." aria-hidden="true" data-contents="...">
 <div class="container">
  <p class="...">By visiting x.gr you agree to the use of cookies.</p>
  <a href="#" class="...">Agree</a>
 </div>
</div>
```

```
{
 "has_accept_btn": 1,
 "accept-btn_cta": "Accept",
 "privacy_text": "By visiting x.gr you agree to the use of cookies.",
 ...
}
```

Fig. 2: A cookie banner before (top) and after (bottom) normalisation

Table 2: Breakdown of the number of websites per country that are included (by source) and excluded (by reason)

	Greece	**UK**	**Combined**
Included from Tranco	3,446	18,768	22,214
Included from country-specific lists	40	634	674
Excluded due to unavailability	−125	−305	−430
Excluded due to `robots.txt`	−204	−3,687	−3,891
Excluded due to Terms of service	−70	−760	−830
Total studied	**3,087**	**14,650**	**17,737**

and 305 UK websites from the dataset as they were not accessible. In total, the initial dataset contained 3,361 Greek and 19,097 UK available websites.

We checked each website to determine whether they allow crawlers. The Robots Exclusion Standard parser yielded 3,157 Greek (93%), and 15,410 UK (69%) websites that allowed crawling. Then the Terms of Service parser determined that 3,087 Greek (91%) and 14,650 UK (65%) websites permitted our study to crawl them. Table 2, summarises the breakdown of our dataset.

OpenWPM. Successfully crawling websites for their cookie banners was possible by extending OpenWPM. In addition to cookie banners, OpenWPM also collected more than 15 M data points in Greece and the UK. This included information about the HTTP Requests and Responses (3.9 M), scripts that a website loads (7.6 M), and cookies stored in a user's web browser (2.3 M).

Viking. Collecting cookie banners for thousands of websites is a highly computing-intensive task, requiring over 24 hours for a complete crawl in Greece for example, even with parallel crawlers. To overcome this limitation we utilised University

Table 3: Comparison of measured cookie banner prevalence rates. Sample sizes are approximates. Ranges indicate two methods of measurement.

Study	Year Conducted	Sample Size		Prevalence	
		UK	GR	UK	GR
Degeling et al. [2]	2018	500	500	67–82%	60–69%
van Eijk et al. [3]	2019	100	100	52%	29%
This work	2020	14,000	3,000	44%	48%

of York's Viking cluster[1], a high-performance computing cluster with 173 nodes, 42 TB of memory, and 7024 Intel cores. While only using a fraction of Viking's resources (128 GB of memory and 32 cores) the crawl was completed in 8 hours for Greece and just over 36 hours for the UK.

4.2 Findings

RQ1: Prevalence depends on sample size. Our findings show that almost half of the websites we surveyed display a cookie banner. More specifically, around 48% of Greek and 44% of the UK websites included a cookie notice.

When comparing our results with previous works of Degeling et al. [2] and van Eijk et al. [3], an interesting pattern emerges. As shown in Table 3, although both van Eijk et al.'s and our data collection were conducted after that of Degeling et al., we report lower prevalence than that of the earlier study. This is at odds with the reasonable expectation that the prevalence of cookie banners does not decrease substantially over time. What can explain this discrepancy is the sample size factor. Our results demonstrate that the observed prevalence depends on the size of the sample. That is, although the observed rates might rise initially as samples are expanded from the top hundred to a few hundred websites in each country, further expansion to a few thousands results in a decrease in observed prevalence rates. Hence, our results show that studies with smaller sample sizes might not provide an accurate representation of the big picture.

Using the additional data collected by OpenWPM, we found that 61% of Greek and 70% of UK websites store at least one third-party cookie on their user's browser. This suggests that around 13% of Greek and 26% UK websites have yet to comply with the GDPR or DPA, respectively, as shown in Fig. 3.

RQ2: Direct opt-outs are rare. The distribution of the number of options cookie banners in our dataset provide is depicted in Fig. 4. As the figure shows, the most prevalent number of options in both countries is two. The mean number of options is 2.1 for Greece and 1.8 for the UK. The median number of options is 2 for both countries. This is in agreement with van Eijk et al.'s finding that the median number of choices in the top 100 popular websites that have a cookie banner in both countries was two [3].

[1] See `www.york.ac.uk/it-services/research-computing/viking-cluster`

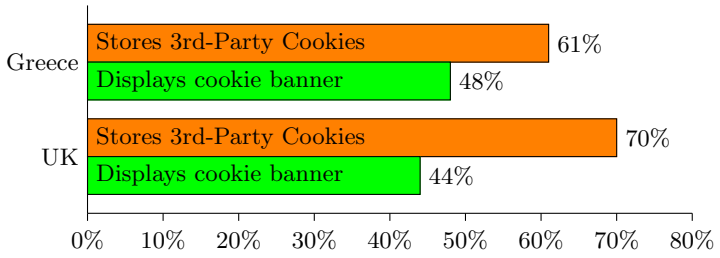

Fig. 3: Websites that store 3rd-party cookies and display a cookie banner.

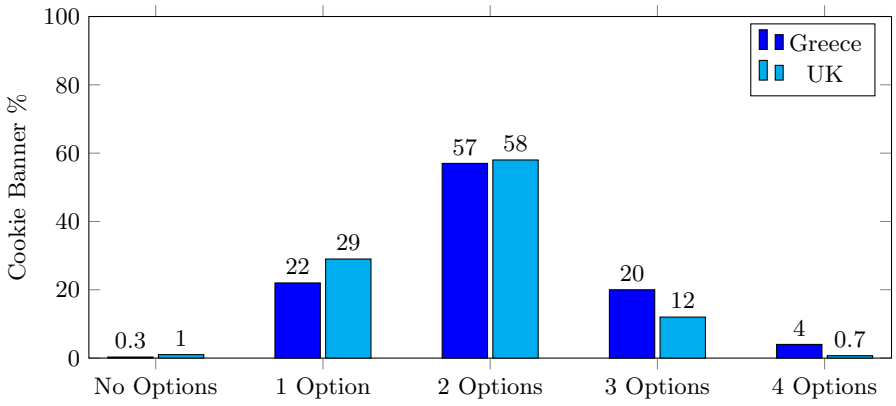

Fig. 4: Distributions of number of cookie banner options in Greece and the UK.

Worryingly, we can see in Fig. 4 that a considerable proportion of cookie banners provide either no option or only one option to the user. This prompts us to look into the distribution of the four categories of privacy options in the cookie banners. The results are depicted in Fig. 5. As the figure shows, although Affirmative options are quite ubiquitous in cookie banners in both countries (Greece: 95%, UK: 88%), Negative options are quite rare (Greece: 20%, UK: 6%). In the upcoming sections we will look further into the exact combinations of options provided by the websites to be able to draw further conclusions.

RQ3: Most cookie banners nudge towards privacy-intrusive choices. Considering all the ways the four categories of options that we have discussed may appear in a cookie banner results in 16 possible combinations. We depict the distributions of all these 16 option combinations in both countries in Table 4. The combinations are coded with abbreviations in the table, e.g. `A-M-` stands for the combinations in which at least an Affirmative choice is present, Negative absent, Managerial present, and Informational absent.

As Table 4 shows, by far the most prevalent combination is that of Affirmative and Managerial options (i.e. `A-M-`), with other combinations including an

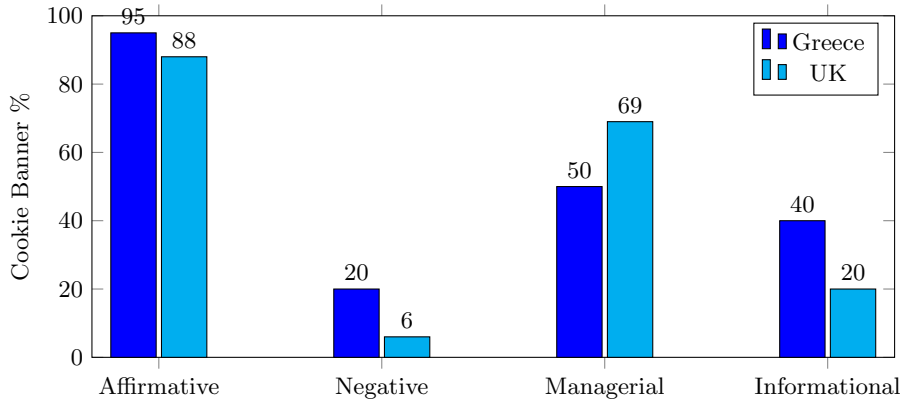

Fig. 5: The proportion of cookie banners in Greece and the UK providing each type of option.

Affirmative option but excluding a Negative option (i.e. `A-MI`, `A--I`, and `A---`) following in terms of prevalence in both countries. This shows that at least 75% of cookie banners in Greece and 82% in the UK explicitly nudge their users towards accepting cookies.

Going beyond nudging, as Nouwens et al. argue [13], implicit consent and reject not being as easy as accept are both violations of GDPR and DPA. Let us now consider the 16 combinations against these two criteria.

For explicit consent, one requires at least an Affirmative or a Managerial option to be present so that the user can register their consent explicitly through one of these options. Hence, all combinations without any of these two options (i.e. `-N-I`, `-N--`, `---I`, and `----`) represent cookie banners that are violating this criterion. Hence, our results show that at least 16 Greek and 129 UK websites are non-compliant with GDPR and DPA since they do not provide the means for their users to register their explicit consent to the use of cookies. These constitute around 1% of Greek and 2% of UK websites with cookie banners.

The proportions of websites not providing an explicit consent option discussed above are large under-estimations since consent is not necessarily explicit in other combinations. More specifically, in our Affirmative category, apart from terms such as "accept" and "I agree" that clearly indicate consent, there are many other terms with less clear meaning such as "close", "continue", and "dismiss". These less clear terms roughly constitute around one sixth of all of the observed Affirmative options. We do not believe that such terms are sufficient to indicate explicit consent and hence estimate the level of non-compliance in terms of explicit consent to be around 15%.

The situation is much worse if the relative ease of Affirmative and Negative options are considered. Any combination with an Affirmative choice but without a Negative one (i.e. `A-MI`, `A-M-`, `A--I`, and `A---`) clearly does not provide a negative option as easily accessible as an Affirmative one. Furthermore, any

Table 4: Distribution of the 16 combinations of Privacy Options in Greece and the UK, highlighting those that directly violate the GDPR and the Data Protection Act 2018. `A`: Affirmative, `N`: Negative, `M`: Managerial, `I`: Informational.

Combination	GR	UK	Consent Explicit	Accept as easy as Reject
`ANMI`	4%	1%		
`ANM-`	5%	3%		
`AN-I`	10%	<1%		
`AN--`	2%	1%		
`A-MI`	5%	8%		No
`A-M-`	32%	47%		No
`A--I`	21%	8%		No
`A---`	17%	19%		No
`-NMI`	<1%	0%		
`-NM-`	<1%	<1%		
`-N-I`	0%	0%	No	
`-N--`	0%	0%	No	
`--MI`	<1%	1%		
`--M-`	4%	9%		
`---I`	1%	1%	No	No
`----`	<1%	1%	No	No

cookie banner that only includes an Informative option or no option (i.e. `---I` and `----`) is defaulting on acceptance of cookies if the user navigates away from the cookie banner to interact with the website, hence not providing any means for the user to register their lack of consent. Therefore, all of these combinations do not satisfy the criterion either. This means that overall, our results demonstrate that at least 76% of Greek and 84% of UK cookie banners violate the GDPR and DPA in that they do not provide their users with a Negative option as easily accessible as an Affirmative one.

Degeling et al. [2] report their observations of a several types of cookie banners of top 500 websites in Greece and the UK. Three of their categories can be roughly comparable to collections of combinations we report. Cookie banners with "no options" in their work roughly correspond to combinations with neither an Affirmative nor a Negative option (i.e. `--??` where `?` is a wildcard). They report around 20% and 40% for this category (estimated from [2, Figure 5(a)]) compared to our 5% and 12% respectively for Greece and the UK. Cookie banners with "confirmation only" in their work roughly correspond to combinations with an Affirmative but not a Negative option (i.e. `A-??`). They report around 65% and 35% for this category (estimated) compared to our 75% and 82%. Cookie banners with a "binary" choice in their work roughly correspond to combinations with both an Affirmative and a Negative option (i.e. `AN??`). They report around 4% and 5% for this category (estimated) compared to our 20% and 5%.

These comparisons show that observed practices may substantially vary between observations of smaller and larger sample sizes.

Limiting their attention to cookie banners provided by the 5 most popular CMPs in the UK, Nouwens et al. found around 75% violating the "reject as easy as accept" criterion [13]. Our analysis gives the rate of at least 84% for the violation of this criterion showing that the situation is much worse when a larger set of websites are considered.

In addition to privacy options, cookie banners usually contain a concise textual description as well. The text's primary function is to inform users why cookies are used and how they may affect them. This text is usually considerably shorter compared to the full Privacy Policy of the website. Examples of such texts can be seen in Fig. 1. The average length of cookie banner texts in Greek websites was 66 words, slightly longer than the UK average of 52 words.

Employing the term frequency–inverse document frequency (TF-IDF) formula to identify the most prominent terms in the cookie banner text corpus, we found that the most prominent terms in Greece and the UK are quite similar and dominated by terms with an apparent positive connotation such as "best"/"better"/"$\kappa\alpha\lambda\acute{\upsilon}\tau\varepsilon\rho\eta$", "ensure", and "experience"/"$\varepsilon\mu\pi\varepsilon\iota\rho\acute{\iota}\alpha$". In fact, none of the top 50 prominent terms in either country (available from the repository) appear to have a negative connotation, whereas terms with a positive connotation such as "improve"/"$\beta\varepsilon\lambda\tau\iota\acute{\omega}\sigma\varepsilon\iota$" and "enhance" constitute a considerable proportion of the list of terms.

To get a more comprehensive view of the connotations relayed by cookie banner texts in the UK, we performed an automated sentiment analysis of the words used in all UK banner texts using NRCLex [1]. The analysis found that a generally positive emotional affect was present in around 80% of the banner texts, whereas a generally negative affect was present in only around 14%. Besides, an overwhelming majority (of more than 9 in 10) of the texts with a negative affect also had a positive effect present as well. Looking at more specific emotional affects, trust and joy are among the most prevalent, present in around 66% and 46% of the texts, respectively. The prevalence of general and specific emotional affects is shown in Fig. 6.

The automated term prominence and sentiment analyses above suggest that websites tend to give a one-sided description of cookie usage, namely that it enhances browsing experience, conveniently leaving out that cookies can be used for tracking. This is in line with previous manual analyses of smaller samples, e.g. that of Utz et al. [14], that found similar biases in cookie banner texts.

RQ4: Managing trackers is more prevalent than opting out. We aimed to determine whether websites allow their visitors to manage their privacy settings from the cookie notice. Our results show that Managerial options in cookie banners are significantly more prevalent compared to the Negative ones (See Fig. 5). More specifically, 59% of Greek and 69% of UK cookie banners offer a Managerial option compared to 20% and 6%, respectively, with a Negative one.

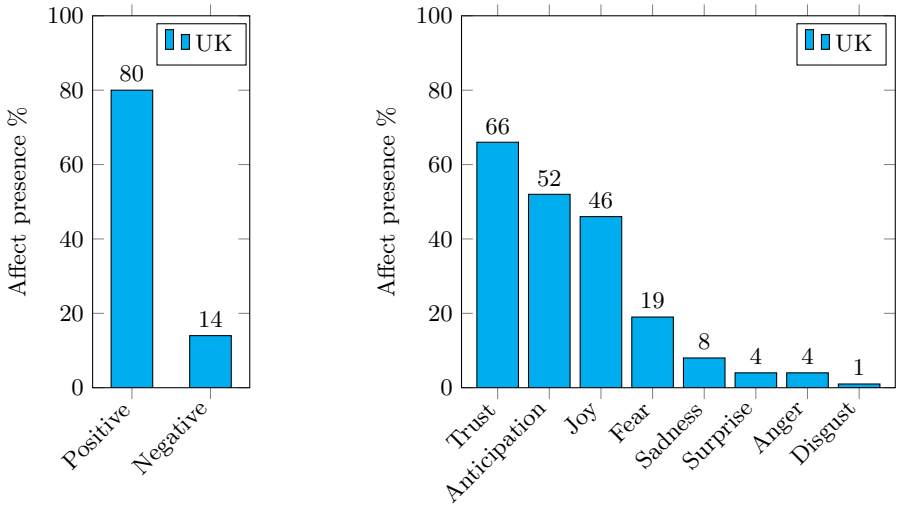

Fig. 6: Distributions of the emotional affects in the UK cookie banner text corpus

Hence, users in both countries are several times more likely to be given an option to manage their cookies than an option to decline them.

RQ5: Users in both countries lack real choice, but practices vary. The results discussed in the previous sections show that Greek and UK users face a largely similar landscape in terms of the prevalence of cookie banners, widespread deployment of third-party cookies, and rampant use of nudging and lack of consent to cookies being much harder to register than consent. However, there are some notable differences between the two countries. With respect to using third-party cookies but not showing a cookie banner at all, the proportion of websites that are non-compliant with regulations in the UK is almost double that in Greece. Besides, cookie banner in Greece tend to provide slightly higher number of options.

Looking at the option types, affirmative options are prevalent in both, but negative options are much scarcer in the UK. From the other two types, managerial options are more prevalent in the UK and informational ones more in Greece. Looking at the option combinations, most have similar prevalence with three exceptions: although `AN-I` can be found in around 10% of banners in Greece, it is very rare in UK; similarly, `A--I` has a much higher share in Greece (21%) than in the UK (8%); on the other hand, `A-M-` is found in almost half of the cookie banners in the UK, but only in about a third in Greece.

5 Conclusion

We set out to conduct the most comprehensive study of cookie banners in the UK and Greece to date in the hope that a more thorough understanding of the cookie banner landscape in the two countries is beneficial for a rage of stakeholders including users, privacy-enhancing technology developers, and policymakers. By extending OpenWPM to detect and store cookie banners, over 17,000 websites were crawled and more than 7,000 cookie banners were collected.

Our results show that although around half of the websites in our dataset display a cookie notice, a substantial proportion do not show one even though they use third-party cookies. Furthermore, websites make it extremely difficult for users to opt-out from tracking with only a minority offering a direct opt-out option. Our analysis also suggests that websites present cookies as devices that improve browsing experience for the user while the negative aspects of tracking tend to be downplayed. Hence, we find clear evidence of websites nudging visitors towards privacy-intrusive choices and violating regulations.

Although in many cases our results agree with previous studies considering smaller samples, we also found that in some cases, e.g. prevalence of cookie banners and those providing specific options, our observations significantly differ from previous reported values. Hence, we hope that our work provides a more holistic view of the landscape of cookie banners in the two countries.

Future work directions include more comprehensive studies of the cookie banner landscape for other countries (for which our code is available and can be reused), a more detailed analysis and classification of varying cookie banner practices in specific subsets of the dataset, e.g., in different industries, and further analyses of the cookie banner text corpus.

References

1. Bailey, M.M.: NRCLex (2019), GitHub Repository, available online at `https://github.com/metalcorebear/NRCLex`
2. Degeling, M., Utz, C., Lentzsch, C., Hosseini, H., Schaub, F., Holz, T.: We value your privacy ... now take some cookies: Measuring the GDPR's impact on web privacy. In: NDSS. The Internet Society (2019)
3. van Eijk, R., Asghari, H., Winter, P., Narayanan, A.: The impact of user location on cookie notices (inside and outside of the european union). In: Workshop on Technology and Consumer Protection (ConPro'19) (2019)
4. Englehardt, S., Narayanan, A.: Online tracking: A 1-million-site measurement and analysis. In: Proceedings of ACM CCS 2016 (2016)
5. European Commission: Digital Economy and Society Index (DESI) Report 2020: Use of internet services (2020)
6. European Data Protection Board: Guidelines 05/2020 on consent under Regulation 2016/679. Version 1.1 (2020)
7. European Union: Directive 2002/58/EC of the European Parliament and of the Council. Official Journal of the European Communities: L 201/37 (2002)
8. Habib, H., Zou, Y., Jannu, A., Sridhar, N., Swoopes, C., Acquisti, A., Cranor, L.F., Sadeh, N., Schaub, F.: An empirical analysis of data deletion and opt-out choices on 150 websites. In: Symposium on Usable Privacy and Security (SOUPS) (2019)

9. Kulyk, O., Hilt, A., Gerber, N., Volkamer, M.: This website uses cookies: Users' perceptions and reactions to the cookie disclaimer. In: European Workshop on Usable Security (EuroUSEC) (2018)
10. Le Pochat, V., Van Goethem, T., Tajalizadehkhoob, S., Korczyński, M., Joosen, W.: Tranco: A research-oriented top sites ranking hardened against manipulation. In: Network and Distributed System Security Symposium. NDSS (2019)
11. Matte, C., Bielova, N., Santos, C.: Do cookie banners respect my choice?: Measuring legal compliance of banners from IAB Europe's transparency and consent framework. In: IEEE Symposium on Security & Privacy. pp. 791–809. IEEE (2020)
12. Norwegian Consumer Council: Deceived by design, how tech companies use dark patterns to discourage us from exercising our rights to privacy. Norwegian Consumer Council Report (2018)
13. Nouwens, M., Liccardi, I., Veale, M., Karger, D., Kagal, L.: Dark patterns after the GDPR: Scraping consent pop-ups and demonstrating their influence. In: CHI Conference on Human Factors in Computing Systems. pp. 1–13 (2020)
14. Utz, C., Degeling, M., Fahl, S., Schaub, F., Holz, T.: (Un)informed Consent: Studying GDPR Consent Notices in the Field. In: ACM SIGSAC Conference on Computer and Communications Security. pp. 973–990 (2019)